Science
AAAS

Second-Order Selection for Evolvability in a Large *Escherichia coli* Population

Robert J. Woods, *et al.*

Science **331**, 1433 (2011);
DOI: 10.1126/science.1198914

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Second-Order Selection for Evolvability in a Large *Escherichia coli* Population

Robert J. Woods,[1]*† Jeffrey E. Barrick,[2]*‡¶ Tim F. Cooper,[3] Utpala Shrestha,[3] Mark R. Kauth,[2] Richard E. Lenski[1,2]¶

In theory, competition between asexual lineages can lead to second-order selection for greater evolutionary potential. To test this hypothesis, we revived a frozen population of *Escherichia coli* from a long-term evolution experiment and compared the fitness and ultimate fates of four genetically distinct clones. Surprisingly, two clones with beneficial mutations that would eventually take over the population had significantly lower competitive fitness than two clones with mutations that later went extinct. By replaying evolution many times from these clones, we showed that the eventual winners likely prevailed because they had greater potential for further adaptation. Genetic interactions that reduce the benefit of certain regulatory mutations in the eventual losers appear to explain, at least in part, why they were outcompeted.

Organisms may vary not only in traits that determine their immediate fitness, but also in their potential to generate better-adapted descendants with new beneficial mutations. Evolutionary potential, or evolvability, can be operationally defined as the expected degree to which a lineage beginning from a particular genotype will increase in fitness after evolving for a certain time in a particular environment (*1*). Evolvability thus reflects a complex probabilistic integration of accessible paths in the fitness landscape influenced by mutation rates, population structure, and epistatic interactions between mutations (*2*–*4*). Experiments with microorganisms have shown that genotypes with elevated mutation rates have greater evolvability under certain conditions (*5*, *6*). The evolutionary potential of microorganisms can also vary when the same mutations have different fitness effects in different genetic backgrounds due to epistatic interactions (*7*–*9*). The extent to which material differences in evolvability of this latter kind—reflecting genetic architecture (*10*), rather than mutation rates—spontaneously arise between lineages within asexual populations and play a role in ongoing evolutionary dynamics is unknown (*11*, *12*).

We found that several genetically distinct subpopulations were already present in a 500-generation sample archived from a now >50,000-generation long-term evolution experiment with the bacterium *Escherichia coli* (*13*, *14*). In particular, we characterized numerous clones sampled at 500, 1000, and 1500 generations for the presence of five previously discovered beneficial mutations (*14*–*18*). Specific mutations in the *rbs* operon and *topA*, *spoT*, and *glmUS* genes fixed in the evolving population between 1000 and 1500 generations, and in *pykF* after 1500 generations (*14*). Two beneficial mutations—the ones affecting *rbs* and *topA*—were already present in many clones sampled at 500 generations (table S1). Thus, we refer to *topA rbs* genotypes sampled at generation 500 as eventual winners (EWs) and to other contemporaneous genotypes as eventual losers (ELs). These genotypes may have other beneficial mutations in these same or other genes, and most such mutations could not have been detected by the assays developed for the known mutations.

We further characterized two clonal isolates of each type: EW1 and EW2, each with the *topA* and *rbs* mutations, EL1 with no known mutations, and EL2 with another mutation we call the *rbs1* mutation. The EW *topA* allele is an amino acid substitution in DNA topoisomerase I that alters chromosomal supercoiling, affects the transcription of many genes, and confers a fitness benefit of ~13% when moved into the ancestral genetic background (*16*). The *rbs* mutations are deletions of different sizes in the ribose utilization operon that occur with high frequency and cause 1 to 2% fitness gains (*15*). Competition experiments against the ancestral strain showed that these four representative clones were 13 to 23% more fit than the ancestor (Fig. 1). Thus, the ELs, and possibly also the EWs, had other beneficial mutations, in addition to those that could be identified by our initial genotyping.

One would naïvely expect that the EWs were better adapted than the ELs at 500 generations, but the opposite was true (Fig. 1). Direct competition experiments also showed that the two representative EW clones were at a significant fitness disadvantage relative to the two representative EL clones (fig. S1). In fact, if the fitness deficit of the EWs (−6.3%) had remained constant, they would have gone extinct in another ~350 generations (*19*). We found no evidence of negative frequency-dependent interactions (*20*) between EW and EL strains that might have stabilized their continued coexistence in the long-term population (*19*).

How did descendants of the EWs prevail over EL lineages despite their fitness deficit? The EW-derived lineage may have simply been "lucky" in this one instance of evolution; that is, they might have stochastically gained highly beneficial mutations that allowed them to overtake the EL subpopulations before they were driven extinct. Alternatively, the EW genotypes may have had a greater potential for further adaptation, such that they would reproducibly give rise to higher-fitness descendants and outcompete EL lineages before they were lost. To distinguish between these two hypotheses, we "replayed" evolution by initiating 10 replicate experimental populations from each clone isolated at 500 generations (EW1, EW2, EL1, and EL2). Each population was propagated independently under the same conditions as were used in the long-term evolution experiment for 883 generations, approximately as long as ELs and EWs coexisted in the original population.

To follow the evolutionary dynamics in more detail, we conducted these evolution experiments in a neutral marker divergence format (*9*, *21*, *22*). A variant of each of the four *E. coli* test strains was constructed wherein the state of a readily scored phenotypic marker (Ara), which has no effect on fitness under these culture conditions, was altered by a specific point mutation (*13*). Each experimental population was then started by mixing approximately equal numbers of the original test strain (Ara^-) and the strain with the changed marker (Ara^+), with each type grown separately from a single colony to ensure that there was essentially no initial genetic variation within a population and no shared history between independent replicates. Tracking the frequency of this genetic marker in these 40 populations over time allowed us to visually follow and quantitatively analyze the first beneficial mutations that swept to high frequency within each population.

The trajectory for the Ara^-/Ara^+ ratio in each population eventually diverged from its starting level, as bacteria with beneficial mutations linked to one marker state arose and outcompeted their ancestors and competitors with less-beneficial

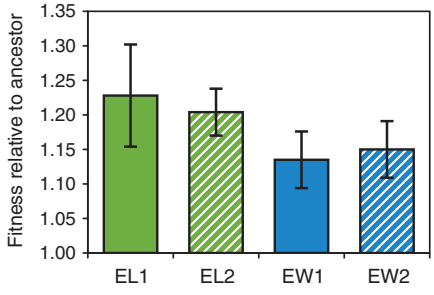

Fig. 1. Fitness of two EW and two EL clones relative to the ancestor of the *E. coli* long-term evolution experiment. Error bars are 95% confidence intervals. All four clones were significantly more fit than the ancestor. Surprisingly, the EL clones were more fit than the EW clones, both as shown here and in direct competition with one another (fig. S1).

[1]Department of Zoology, Michigan State University, East Lansing, MI 48824, USA. [2]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA. [3]Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA.

*These authors contributed equally to this work.
†Present address: Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA.
‡Present address: Department of Chemistry and Biochemistry, Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX 78712, USA.
¶To whom correspondence should be addressed. E-mail: jbarrick@cm.utexas.edu (J.E.B.); lenski@msu.edu (R.E.L.)

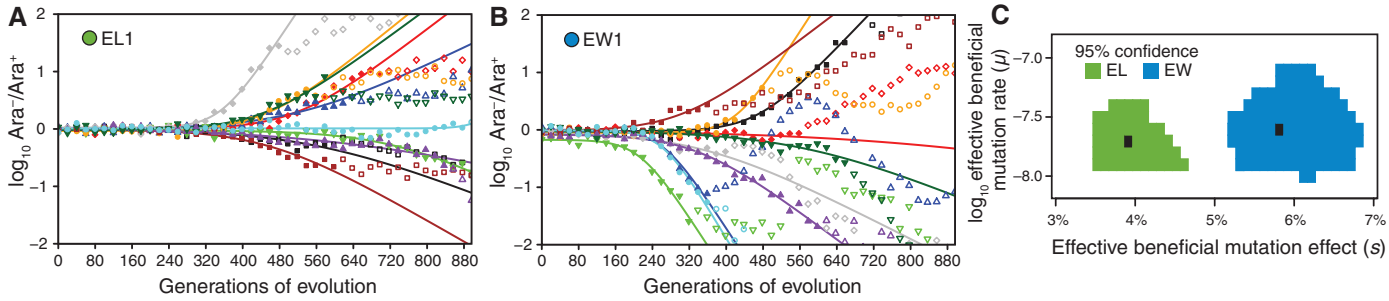

Fig. 2. Replay evolution experiments to measure the evolvability of the four representative 500-generation clones. (**A** and **B**) The frequencies of Ara$^-$ and Ara$^+$ versions of each test strain, initially mixed equally, were recorded at regular intervals (symbols) during 883 generations of evolution under the same conditions as were used in the long-term experiment. Marker trajectories for the replay populations initiated from EL1 and EW1 clones are shown (10 replicates each). Shifts in the Ara$^-$/Ara$^+$ ratio occur when new beneficial mutations linked to one background arise and increase in frequency within the population. Fitting the replicate marker trajectories (lines and solid symbols) until they deviate significantly from an exponential model (open symbols) provides a distribution of empirical shape parameters for the initial divergence. (**C**) Effective mutation rates (μ) and fitness effects (*s*) for the first beneficial mutations to sweep to high frequency in a given genetic background were inferred by comparing experimental divergence parameters with those from simulated marker trajectories. Black rectangles represent maximum-likelihood estimates. Representative EW and EL isolates were grouped for this analysis (*19*). Figure S2 shows data for EL2 and EW2 populations and other steps in the statistical analysis.

mutations (Fig. 2, A and B, and fig. S2, A and B). The shape of the initial divergence of the family of curves generated from evolutionary replicates of the same clone reflects its local fitness landscape. In particular, a simple model that assumes one category of beneficial mutation, with an effective mutation rate (μ) and fitness benefit (s), reproduces the salient features of these dynamics (*21*), provided that it includes competition between lineages with alternative beneficial mutations, i.e., clonal interference (*2*, *23*, *24*).

By performing population genetic simulations to generate families of curves for many μ and s values (*19*), we determined the parameter combinations that agreed best with the experimental curves (Fig. 2C). The replicate EW marker ratio trajectories diverged earlier and more steeply than the EL trajectories, and the effective size of the first beneficial mutations to sweep to high frequency was significantly larger for the EWs. However, by itself, this first mutation was not sufficient for the EWs to overtake the ELs. From the 6.3% initial fitness deficit of the EWs and the 1.8% larger effect size of their first beneficial mutations, we calculate that, on average, the EWs would still be ~4.5% less fit than the ELs after the first adaptive step for each type (Fig. 3).

To compare evolvability on a time scale that allows a lineage to accumulate multiple beneficial mutations, we isolated a random clone from each replicate population at the 883-generation endpoint. This evolved clone was either Ara$^-$ or Ara$^+$. We performed head-to-head competitions of every EL-EW pair with opposite Ara marker states to determine their relative fitness (fig. S3). We found that, on this time scale, the EWs overcame their initial fitness deficit and evolved to higher fitness than the ELs by ~2.1%, on average (Fig. 3). Thus, the EWs evidently prevailed in the original long-term population because they had greater evolvability. On average, they achieved higher fitness than the ELs after each type had equal time to evolve by multiple beneficial-mutation steps from its starting point in the fitness landscape. We stress, however, that this result is necessarily probabilistic in nature. Not every evolved EW clone was able to outcompete every evolved EL clone.

What is the genetic basis of this difference in evolvability? There are three possibilities. First, the EW genetic background may have interacted more favorably with certain potential beneficial mutations than the ancestral background with those same mutations (positive epistasis), thereby opening up additional pathways for adaptive evolution. Conversely, mutations in the ELs may have reduced the effects of otherwise beneficial mutations (negative epistasis), thereby closing off some pathways for adaptation. Finally, a mutation in the EWs may have caused an elevated mutation rate relative to that of the ELs that would allow the EWs to access rarer, more beneficial mutations.

To distinguish the salient genetic differences between the EWs and ELs, we resequenced the genomes of eight evolved *E. coli* isolates from generation 883 of the replay experiments (*19*). We chose two strains descended from each of the four 500-generation clones, so that we could reconstruct what mutations were present in the original isolates as well as sample mutations that occurred in their descendants (Fig. 4A). We found that the EWs shared only the two known mutations (*topA* and *rbs*) and that both ELs had two previously unknown base substitutions (*topA1* and *fadR*). The EL *topA1* mutation alters the amino acid (isoleucine-34 to serine) directly adjacent to the one changed by the EW *topA* allele (histidine-33 to tyrosine). FadR is a regulator of fatty acid and acetate metabolism (*25*), and the effects of this EL mutation are unknown.

From two to five mutations accumulated during the 883-generation replay experiment in the eight independently evolved isolates (tables S2 to S9). There was no evidence that EWs had an elevated mutation rate that might have contributed to their greater evolvability. The number of replay-phase mutations in the four evolved EWs (16 total) was essentially identical to that in the four evolved ELs (15 total). The only mutation in the original

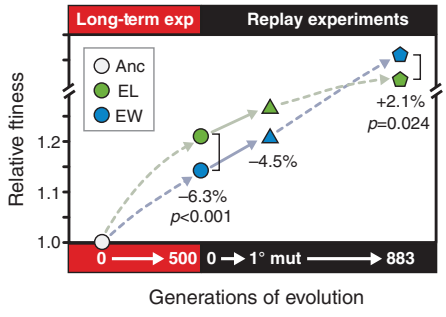

Fig. 3. Greater evolvability of EWs allows them to reproducibly overtake ELs. Two representative EW clones from generation 500 of the long-term evolution experiment were initially at a significant fitness disadvantage relative to two contemporary EL clones (circles). The EWs were somewhat closer in fitness to the ELs, but still lagged behind on average, after the first beneficial mutations swept to high frequency in the replay evolution experiments (triangles), as determined by the marker trajectory divergence analysis. After 883 generations, the representative EWs evolved to a higher fitness on average than the ELs in the replay populations (pentagons). Percentage differences in fitness are for pooled EWs versus ELs at the highlighted time point, and *P*-values indicate whether this difference was significant (*19*). Arrows represent presumptive mutational steps, with dashes indicating that the exact number of mutations may vary. The *y* axis is unlabeled for the final 883-generation replay isolates because their fitness was measured with respect to each other, not relative to the ancestor.

500-generation clones in a gene related to DNA repair or replication (*uvrB*) was shared by the EL1-derived strains, contrary to the expectation if a change in mutation rate drove the difference in evolvability.

It is very likely that most of the 31 observed mutations that occurred during the replay experiments are beneficial. The rate of genomic change due to adaptive evolution greatly exceeds the rate of neutral drift in this system, and mutations in

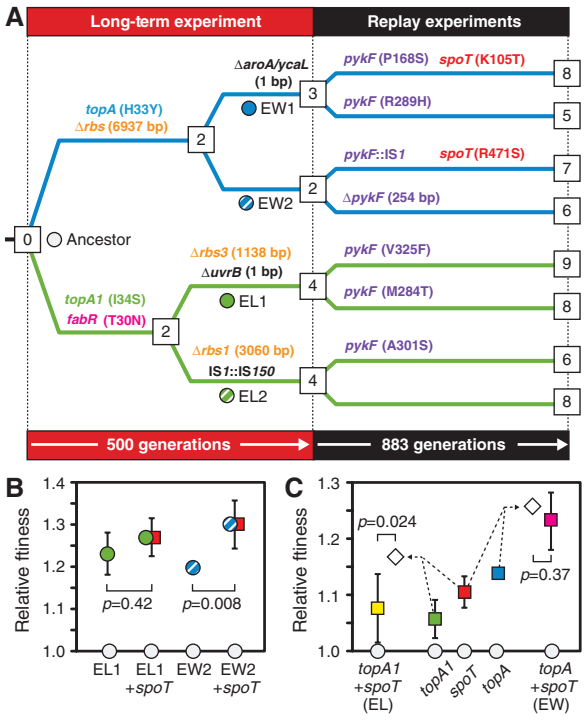

Fig. 4. (**A**) Mutations identified by whole-genome resequencing of endpoint *E. coli* clones from the replay evolution experiments and inferred by parsimony in their EW and EL ancestors. Numbers in squares indicate how many mutations accumulated relative to the original long-term ancestor with mutations in key genes labeled. See tables S2 to S9 for complete lists of all mutations. (**B**) Fitness effects of adding the *spoT* mutation that fixed during the long-term experiment to the EL1 and EW2 genetic backgrounds, measured relative to the ancestral strain. Error bars are 95% confidence limits (hidden by symbols in some cases). *P*-values indicate the significance of the hypothesis that addition of the *spoT* mutation caused a fitness difference. (**C**) Fitness effects of the long-term *spoT*, EW *topA*, and EL *topA1* mutations alone and in combination in the ancestral genetic background. Dashed lines converging on empty diamonds show the fitness predicted for each *spoT* and *topA* allele combination given independent multiplicative effects. *P*-values are for the hypothesis of no epistatic interactions under a multiplicative model (*31*). Error bars are 95% confidence limits.

some of the same genes, operons, and pathways have been found in other isolates from the long-term experiment (*14*, *26*). Two genes independently evolved mutations in more than one of the sequenced 883-generation EL- and EW-derived strains (Fig. 4A). Seven of the eight sequenced clones from the replays evolved mutations in *pykF*, which encodes the metabolic enzyme pyruvate kinase. In the original long-term population, *pykF* mutations were detected in EWs by 1500 generations (table S1) and in *rbs1* ELs by 1000 generations (*19*). The long-term EW *pykF* mutation is highly beneficial in this environment (*14*), and all 12 long-term populations substituted *pykF* mutations by 20,000 generations (*27*). Two of the sequenced replay clones had point mutations in *spoT*, which encodes a bifunctional (p)ppGpp synthesis and degradation enzyme that is a global regulator of gene expression. The next mutation found in the long-term EW lineage, after *rbs* and *topA*, was also a *spoT* base substitution (table S1). This *spoT* mutation has been shown to affect the transcription of numerous genes and to confer a fitness benefit of ~9% when moved into the ancestral strain background (*17*, *28*).

We observed *pykF* mutations in both EW- and EL-derived strains. However, both *spoT* mutations arose in EW-derived lines (Fig. 4A). This association with genetic background raised the possibility that *spoT* mutations might be involved in epistatic interactions that underlie differences in EW-EL evolvability, especially given the potential for widespread pleiotropic effects caused by mutations in this global regulator (*29*). To increase the statistical power for detecting an association, we sequenced the complete *spoT* reading frame in all 40 EW and EL endpoint clones isolated from the replay experiments. We found *spoT* mutations in 6 of the 20 evolved EW clones (table S10). Notably, we found no *spoT* mutations in the 20 EL clones, a difference unlikely by chance (two-tailed Fisher's exact test, $P = 0.0202$).

To test directly for epistatic interactions, we measured the fitness effects of adding the *spoT* allele that arose during the long-term experiment to the EW2 and EL1 strain backgrounds. We found that this *spoT* allele confers a large fitness benefit in the EW background, but has no significant effect in the EL background (Fig. 4B). Because changes in chromosomal supercoiling also may have widespread pleiotropic effects, it seemed likely that interactions with the *topA* alleles specific to the EW and EL backgrounds might explain these epistatic effects. To test this hypothesis, we constructed otherwise isogenic strains carrying only the EW *topA* or EL *topA1* alleles on the ancestral background and then measured the fitness effects of adding the *spoT* mutation that arose during the long-term experiment to each strain (Fig. 4C). The EL *topA1* allele is beneficial on its own, though less so than the EW *topA* mutation. The *spoT* mutation is also highly beneficial on its own and has essentially the same fitness effect in the presence of the EW *topA* allele. By contrast, this *spoT* mutation is neutral, or at least much less beneficial, in combination with the EL *topA1* mutation.

These results therefore support the hypothesis of negative epistasis between EL mutations and later mutations that arise in *spoT*, and they contradict the hypothesis of positive epistasis between the EW and that *spoT* allele. Highly beneficial *spoT* mutations are evidently at the leading edge of the adaptive mutations that are accessible to the EW subpopulation, which is why they often evolve in this background. The alternative *topA1* mutation that arose in the EL subpopulation evidently renders those *spoT* mutations neutral, leaving other less-beneficial mutations the best available. In essence, the ELs followed a trajectory in the fitness landscape that allowed more rapid improvement early on, but which shut the door on at least one important avenue for further improvement. By contrast, the EWs followed a path that did not preclude this option, giving them a better than otherwise expected chance of overtaking the ELs. Because *spoT* mutations evolved in only 6 of the 20 EW replays, and because it took multiple mutations for the EWs to overtake the ELs, it is likely that epistatic interactions with other beneficial mutations also contributed to their differences in evolutionary potential.

We have demonstrated in detail a case in which epistatic interactions between beneficial mutations caused differences in bacterial evolvability that appear to have played a pivotal role in the evolution of a population. Similar cases are expected in any population of asexual organisms that evolve on a rugged fitness landscape with substantial epistasis, as long as the population is large enough that multiple beneficial mutations accumulate in contending lineages before any one mutation can sweep to fixation (*2*, *11*, *22*–*24*, *30*). This scenario thus provides a general mechanism for the evolution of evolvability by "second-order selection" (*3*) on genetic architecture. Our results also suggest that studying the interactions among regulatory networks could lead to a deeper understanding of how genetic changes in those networks might either promote or impede evolvability at a systems level.

References and Notes

1. M. Pigliucci, *Nat. Rev. Genet.* **9**, 75 (2008).
2. C. A. Fogle, J. L. Nagle, M. M. Desai, *Genetics* **180**, 2163 (2008).
3. O. Tenaillon, B. Toupance, H. Le Nagard, F. Taddei, B. Godelle, *Genetics* **152**, 485 (1999).
4. D. M. Weinreich, R. A. Watson, L. Chao, *Evolution* **59**, 1165 (2005).
5. L. Chao, E. C. Cox, *Evolution* **37**, 125 (1983).
6. J. A. G. M. de Visser, C. W. Zeyl, P. J. Gerrish, J. L. Blanchard, R. E. Lenski, *Science* **283**, 404 (1999).
7. C. L. Burch, L. Chao, *Nature* **406**, 625 (2000).
8. R. C. McBride, C. B. Ogbunugafor, P. E. Turner, *BMC Evol. Biol.* **8**, 231 (2008).
9. J. E. Barrick, M. R. Kauth, C. C. Strelioff, R. E. Lenski, *Mol. Biol. Evol.* **27**, 1338 (2010).
10. T. F. Hansen, *Annu. Rev. Ecol. Evol. Syst.* **37**, 123 (2006).
11. P. D. Sniegowski, P. J. Gerrish, *Philos. Trans. R. Soc. London B Biol. Sci.* **365**, 1255 (2010).
12. N. Colegrave, S. Collins, *Heredity* **100**, 464 (2008).
13. R. E. Lenski, M. R. Rose, S. C. Simpson, S. C. Tadler, *Am. Nat.* **138**, 1315 (1991).
14. J. E. Barrick *et al.*, *Nature* **461**, 1243 (2009).
15. V. S. Cooper, D. Schneider, M. Blot, R. E. Lenski, *J. Bacteriol.* **183**, 2834 (2001).
16. E. Crozat, N. Philippe, R. E. Lenski, J. Geiselmann, D. Schneider, *Genetics* **169**, 523 (2005).

17. T. F. Cooper, D. E. Rozen, R. E. Lenski, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1072 (2003).
18. M. T. Stanek, T. F. Cooper, R. E. Lenski, *BMC Evol. Biol.* **9**, 302 (2009).
19. Materials and methods are available as supporting material on *Science* Online.
20. S. F. Elena, R. E. Lenski, *Evolution* **51**, 1058 (1997).
21. M. Hegreness, N. Shoresh, D. Hartl, R. Kishony, *Science* **311**, 1615 (2006).
22. K. C. Kao, G. Sherlock, *Nat. Genet.* **40**, 1499 (2008).
23. P. J. Gerrish, R. E. Lenski, *Genetica* **102-103**, 127 (1998).
24. S. C. Park, J. Krug, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18135 (2007).
25. Y. Xu, R. J. Heath, Z. Li, C. O. Rock, S. W. White, *J. Biol. Chem.* **276**, 17373 (2001).
26. J. E. Barrick, R. E. Lenski, *Cold Spring Harbor Symp. Quant. Biol.* **74**, 119 (2009).
27. R. Woods, D. Schneider, C. L. Winkworth, M. A. Riley, R. E. Lenski, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9107 (2006).
28. T. F. Cooper, S. K. Remold, R. E. Lenski, D. Schneider, *PLoS Genet.* **4**, e35 (2008).
29. N. Philippe, E. Crozat, R. E. Lenski, D. Schneider, *Bioessays* **29**, 846 (2007).
30. D. E. Rozen, M. G. Habets, A. Handel, J. A. de Visser, *PLoS ONE* **3**, e1715 (2008).
31. J. da Silva, M. Coetzer, R. Nedellec, C. Pastore, D. E. Mosier, *Genetics* **185**, 293 (2010).
32. This work was supported by the NSF (DEB1019989 to R.E.L.), Defense Advanced Research Projects Agency (HR0011-09-1-0055 to R.E.L. and T.F.C), the NIH (K99GM087550 to J.E.B.), and the McDonnell Foundation (220020174 to T.F.C). We thank N. Hajela for laboratory assistance, staff at the Michigan State University (MSU) Research Technology Support Facility for assistance with genome sequencing, and the MSU High-Performance Computing Cluster for computational support. Genome sequencing data have been deposited in the NCBI Short Read Archive (SRA024331).

Supporting Online Material

www.sciencemag.org/cgi/content/full/331/6023/1433/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 to S10
References and Notes

12 October 2010; accepted 1 February 2011
10.1126/science.1198914

Independent and Parallel Recruitment of Preexisting Mechanisms Underlying C_4 Photosynthesis

Naomi J. Brown, Christine A. Newell, Susan Stanley, Jit E. Chen, Abigail J. Perrin, Kaisa Kajala, Julian M. Hibberd*

C_4 photosynthesis allows increased photosynthetic efficiency because carbon dioxide (CO_2) is concentrated around the key enzyme RuBisCO. Leaves of C_4 plants exhibit modified biochemistry, cell biology, and leaf development, but despite this complexity, C_4 photosynthesis has evolved independently in at least 45 lineages of plants. We found that two independent lineages of C_4 plant, whose last common ancestor predates the divergence of monocotyledons and dicotyledons about 180 million years ago, show conserved mechanisms controlling the expression of genes important for release of CO_2 around RuBisCO in bundle sheath (BS) cells. Orthologous genes from monocotyledonous and dicotyledonous C_3 species also contained conserved regulatory elements that conferred BS specificity when placed into C_4 species. We conclude that these conserved functional genetic elements likely facilitated the repeated evolution of C_4 photosynthesis.

Plants use CO_2 as a substrate to make sugars through a complex biochemical process involving the Sun's energy. The key photosynthesis enzyme RuBisCO (ribulose-1,5-bisphosphate carboxylase-oxygenase) fixes CO_2 into a three-carbon molecule, but it does not completely distinguish between O_2 and CO_2 (*1*). Because of this, at least 45 lineages from 19 families of flowering plants have evolved C_4 carbon–concentrating mechanisms that enhance the concentration of CO_2 around RuBisCO and increase the efficiency of photosynthesis by reducing the energy wasted when RuBisCO fixes O_2 (*2*).

All C_4 lineages generate high concentrations of CO_2 around RuBisCO by compartmentalizing photosynthesis, either within or between cells (*3*, *4*), and normally carbon is fixed in mesophyll (M) cells to generate high concentrations of a four-carbon acid. The acid then diffuses through plasmodesmata into adjacent cells, often the cells of the bundle sheath (BS), where decarboxylases release CO_2 around RuBisCO (*3*). In some cases, multiple C_4 acid decarboxylases are used to deliver CO_2 to RuBisCO (*5*, *6*). High activities of these C_4 acid decarboxylases around the veins of C_3 species may have facilitated the polyphyletic evolution of C_4 photosynthesis (*7*, *8*).

The compartmentalization of photosynthesis results from substantial differences in gene expression in C_4 leaves relative to those of C_3 species (*9*, *10*). Genes important for C_4 photosynthesis have been recruited from orthologs present in C_3 species, and in many cases this has involved alterations in spatial patterns of expression, restricting or enhancing the accumulation of proteins to specific cell types (*11*, *12*). For example, in most cases, carbonic anhydrase, phospho*enol*pyruvate carboxylase, and pyruvateorthophosphate dikinase accumulate preferentially in M cells, whereas a C_4 acid decarboxylase and RuBisCO accumulate in BS cells (*13*, *14*). As genes are recruited into C_4 photosynthesis, alterations to the sequence of C_3 genes can lead to cis-elements that allow accumulation of the cognate protein in M or BS cells (*12*). Despite large differences in gene expression between M and BS cells of C_4 leaves (*9*, *10*) and the recruitment of some genes belonging to multigene families into the C_4 pathway, no cis-elements have been identified that regulate multiple genes required for C_4 photosynthesis (*15*–*17*).

NAD-dependent malic enzyme (NAD-ME) accumulates in BS cells of NAD-ME–type C_4 plants, where it releases CO_2 around RuBisCO (*18*). Because NAD-ME is encoded by two genes that generate a heterodimer (*8*), we used it as a model to understand how gene families are recruited into C_4 photosynthesis. We isolated full-length transcript sequences for two *NAD-ME* genes from *Cleome gynandra*, which is the C_4 plant most closely related to the model *Arabidopsis thaliana* (*18*). RNA hybridization in situ showed that *CgNAD-ME1&2* transcripts are specific to BS cells (Fig. 1, A and B). Genomic sequences for the two genes were isolated, and a translational fusion between *CgNAD-ME1* (promoter, exons, and introns) and *uidA* encoding the β-glucuronidase (GUS) reporter was generated. Stable transformants of *C. gynandra* expressing *CgNAD-ME1::uidA* accumulated GUS specifically in BS cells (Fig. 1C). When introns were removed and the *CgNAD-ME1* promoter was replaced by the CaMV 35S promoter that directs expression in both M and BS cells (*17*, *19*), GUS also accumulated specifically in BS cells (Fig. 1D). This result contrasted with the accumulation of GUS in both M and BS cells when the *CgNAD-ME1* gene was not present (Fig. 1E and fig. S1A). These data indicate that sequences in the transcribed region of the gene are sufficient for BS-specific accumulation of CgNAD-ME1.

Microprojectile bombardment of *pCaMV 35S::CgNAD-ME1::uidA* generated GUS specifically in BS cells (Fig. 1G and fig. S1B), in contrast to a *CaMV 35S::uidA* control in which M and BS cells accumulated GUS in roughly equal proportions (Fig. 1F). The full-length coding region of *CgNAD-ME2* under control of the CaMV 35S promoter also resulted in BS-specific accumulation of GUS in *C. gynandra* (Fig. 1H). This suggests that for both *CgNAD-ME* genes the same mechanism generates BS-specific accumulation.

A 3′ deletion series of *CgNAD-ME1* indicated that 240 nucleotides toward the 5′ end were sufficient for BS-specific accumulation of GUS, as was the equivalent region of *CgNAD-ME2* (Fig. 2, A and B, and fig. S1, C to E). This BS specificity was due to defined sequence, because when we

Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK.

*To whom correspondence should be addressed. E-mail: julian.hibberd@plantsci.cam.ac.uk